Journal of Intelligent Systems and Internet of Things
Vol. 10, No. 02, PP. 38-48, 2023
DOI: https://doi.org/10.54216/JISIoT.100204

An Ensemble Learning Approach for detection of Chronic Kidney Disease (CKD)

B. Narasimha Swamy[1], Rajeswari Nakka[2], Aditi Sharma[3*,6], S. Phani Praveen[4], Venkata Nagaraju Thatha[5], Kumar Gautam[7]

[1,4]Department of CSE, PVP Siddhartha Institute of Technology, Vijayawada, India
[2]Department of CSE, Seshadri Rao Gudlavalleru Engineering College, India
[3]Department of Computer Science and Engineering, Symbiosis Institute of Technology, Symbiosis International University, Pune, India
[5]Department of Information Technology, MLR Institute of Technology, Hyderabad
[6]IEEE Senior Member, Symbiosis International University, Pune, India
[7]Gwangju Institute of Science and Technology, South Korea
Emails: swamy_bn@pvpsiddhartha.ac.in; rajeswari.gec@gmail.com; aditi.sharma@ieee.org; phani.0713@gmail.com, nagaraju.thatha@gmail.com; edumonk@ieee.org
*Corresponding Author: aditi.sharma@ieee.org

Abstract

Chronic kidney disease (CKD) is a common and possibly fatal condition affecting billions worldwide. Early detection and accurate diagnosis of CKD are critical for timely intervention and improved patient outcomes. In recent years, machine learning techniques have shown great promise in assisting medical professionals in detecting and diagnosing various diseases. This study aims to develop a novel machine learning (ML) model for detecting CKD using clinical and demographic data. The dataset used in this study comprises a comprehensive collection of patient records, including laboratory test results, medical history, and demographic information. Feature selection is one of the techniques that, combined with the ML approach, select the significant features. Several ML algorithms were implemented to detect CKD in the early stages but identified the issues with existing ML algorithms. The developed models' performance is assessed using precision, accuracy, and recall metrics. Additionally, feature importance analysis is conducted to identify the key factors influencing CKD diagnosis. The strength of the proposed approach shows accurately by identifying the individuals at risk of CKD and distinguishing between different stages of the disease. The dataset used for this research was collected from the UCI repository, which consists of 25 attributes, 550 samples, 400 CKD affected, and 150 standard models. The dataset consists of two folders, training and testing. The training utilizes 1000 samples with detailed patient health conditions. The developed CKD detection model shows promising results, achieving high accuracy of 97.98%. on the test dataset. By leveraging machine learning algorithms, this approach can assist healthcare professionals in making more informed decisions regarding early intervention and personalized treatment plans for patients with CKD. Ultimately, applying machine learning techniques in CKD detection can improve patient outcomes and reduce healthcare costs.

Received: April 11, 2023 Revised: July 04, 2023 Accepted: October 08, 2023

Keywords: Chronic kidney disease (CKD); Machine Learning (ML); Support Vector Machines (SVM); Random Forests; And Neural Networks.

1. Introduction

Chronic kidney disease (CKD) is a long-term condition that damages the kidneys and decreases their ability to filter waste from the blood. CKD can progress over time and eventually lead to kidney failure.

The early stages of CKD often have no symptoms. As CKD progresses, people may experience symptoms such as fatigue, weakness, shortness of breath, high blood pressure, swelling in the feet and ankles, increased urination, especially at night, dark-colored urine, loss of appetite, metallic taste in the mouth, and nausea and vomiting.

The most common causes of CKD are diabetes and high blood pressure. Other causes of CKD include glomerulonephritis (inflammation of the glomeruli, the tiny filtering units in the kidneys), polycystic kidney disease (a genetic condition that causes cysts to form in the kidneys), interstitial nephritis (inflammation of the interstitial tissue in the kidneys), obstructive nephropathy (blockage of the urinary tract), and kidney stones.

There is no cure for CKD, but there are treatments that can help to slow the progression of the disease and prevent complications. Treatment options for CKD include medications to lower blood pressure and cholesterol, medications to control blood sugar levels, medications to reduce protein in the urine, diet and lifestyle changes, such as eating a healthy diet, exercising regularly, and quitting smoking, and dialysis or kidney transplant in severe cases.

The best way to prevent CKD is to control blood pressure and blood sugar levels. Other ways to prevent CKD include eating a healthy diet, exercising regularly, maintaining a healthy weight, quitting smoking, and getting regular checkups and screenings. Early detection and treatment of CKD can help to prevent serious complications.

CKD is a typical situation for the kidneys that slowly reduce the functionality over time. It is an epidemic affecting billions of individuals globally and can result in severe medical conditions such as kidney damage, coronary artery disease, and death [1,2,3]. ML models have shown great potential in healthcare, including the detection and treatment of CKD. One way ML can be used in CKD is by developing a predictive approach that finds individuals with high risk by detecting the disease or progressing to more advanced stages. These models can be trained using data from electronic health records, laboratory tests, and other sources to identify patterns and risk factors that may be missed by traditional diagnostic methods. Another way that machine learning can be used in CKD is to develop models that can predict disease outcomes, such as kidney failure or mortality, based on a human factors and medical history. These models used to guide treatment decisions and improve patient outcomes.

Finally, ML improves the efficiency and accuracy of CKD[4,5] diagnosis and treatment. For example, ML algorithms used to analyze medical images and identify early signs of kidney damage that may be missed by human observers. Machine learning can also be used to optimize treatment protocols and dosages based on a patient's individual characteristics and response to therapy. The ML[6,7,8,9] has the potential to revolutionize the diagnosis, treatment, and management of CKD. By leveraging the power of artificial intelligence and big data, we can develop more accurate and personalized models for CKD prediction, diagnosis, and treatment, ultimately improving patient outcomes and reducing the burden of this widespread disease.

Types of dataset used to detect CKD

Early detection and treatment of CKD are critical. In healthcare and medicine, various datasets are used to develop machine-learning models for CKD detection. Typically, these datasets contain patient information, clinical data, and laboratory test results. The following types of data available for the detection CKD:

- **Clinical Datasets:** These are the most important sources of information for detecting CKD. They include demographics (age, gender), medical history, medications, a family history of CKD, and comorbidities (e.g., hypertension, diabetes) from patients. Symptoms and physical examinations are frequently included in these datasets.
- **Laboratory Test Datasets:** These are required for the diagnosis of CKD. These datasets include data from blood and urine tests such as creatinine levels, glomerular filtration rate (GFR), albuminuria, hematuria, electrolyte levels, and other biomarkers. Anomalies in these test results may indicate CKD.
- **Imaging Datasets:** Imaging datasets include medical images such as ultrasound, computed tomography (CT) scans, and magnetic resonance imaging (MRI). Radiological images can provide important insights into the structure and function of the kidneys, assisting in diagnosing conditions such as renal cysts, tumors, and hydronephrosis, all of which can contribute to CKD.
- **Genomic Datasets:** Genetics may influence CKD susceptibility. Genomic datasets contain information about a patient's genetic makeup and may aid in identifying genetic markers associated with CKD. This data can be used to assess risk and personalize medicine.
- **Longitudinal Datasets:** It follows patients over time, providing a historical view of their CKD progression. These datasets are useful for studying disease evolution, identifying risk factors, and evaluating treatment efficacy.
- Electronic Health Records (EHRs): EHRs consists of detailed patient information such as clinical notes, medication history, and ICD-10 diagnostic codes. EHR datasets are frequently used in retrospective studies and can aid in identifying CKD cases and associated factors.

- **Public Health Datasets:** Health organizations and research institutions collect and distribute CKD-related datasets. Data on CKD prevalence, epidemiological studies, and population health indicators may be included in these datasets.
- **Synthetic Datasets:** When real-world CKD data is too sensitive or limited, synthetic datasets are created using statistical techniques to simulate real-world CKD data. These datasets can be used to develop and test algorithms while protecting patient privacy.

2. Literature Survey

A. Chardonnay et al. [10] introduced a new prediction classifier to detect CKD based on the predicted values. The proposed classifier achieved better results compared with previous classifiers. Z. Chen et al. [11] proposed a risk prediction model that predicts CKD based on early symptoms and also analyzes the functionality of kidneys. P. Chittora et al. [12] introduced the ML model to predict the early stages of CKD. The comparison between seven classifiers was applied to real-time datasets. The proposed model selects the advanced pre-processing step and feature extraction and applies a classifier to classify the CKD-affected and not affected samples. Manzano et al. [13] proposed a new CKD prediction model based on the patient's age. The health conditions of the patient also analyze the kidney functionality. D. Dua et al. [14] introduced the kidney dataset collected from the UCI repository. K. Eroglu et al. [15] proposed a new ML algorithm that can detect and diagnose the CDK in the early stages based on the patient's health conditions. M. Fatima et al. [16] discussed various ML algorithms that can be used to diagnose kidney disease in the disease doubted patient. Debal et al. [17] proposed a new CKD detection model predicting early CKD stages. The proposed approach analyzes the health conditions based on the patient data, such as low blood count, acidic abnormalities, etc. The proposed method combines various ML algorithms that classify the standard and disease-based samples. Finally, the proposed approach achieved the best accuracy for detecting CKD. Kapoor et al. [18] introduced a new ML model that predicts CKD based on the automation system. The proposed method is compared with various ML algorithms. Among all the algorithms, RF achieved a high accuracy in terms of accuracy of 98.56%. Abdel-Fattah et al. [19] proposed the hybrid ML approach that classifies CKD detection with the combination of feature selection techniques called Relief-F and chi-squared to extract the significant features. The proposed approach shows 100% accuracy in CKD detection, but the processing time takes longer compared with existing systems. Qin et al. [20,21] introduced the ML approach that detects CKD based on the morbidity and mortality rate. Detecting and diagnosing CKD early is essential by combining the two ML models, such as LR and RF, with the RNN as the training model. The proposed approach achieved an accuracy of 99.87% after ten iterations.

Reasons for Causing of CKD

CKD is a global public health problem with severe implications for people and medical systems. It is distinguished by a gradual kidney function impairment over time and is frequently asymptomatic until the advanced stages. Recognizing the causes and progression of CKD is critical for effective CKD prevention, early detection, and management. This section aims to shed light on the multifaceted nature of CKD and the primary causes of its effects, emphasizing the importance of addressing this growing health issue. Millions of people worldwide suffer from CKD, significantly damaging healthcare resources. The prevalence of CKD has steadily increased, partly due to population aging and increased risk factors such as diabetes and hypertension. As a result, CKD has become a significant driver of healthcare costs and a leading cause of morbidity and mortality worldwide. One of the most challenging aspects of CKD is its often asymptomatic and silent progression. Many people are unaware of kidney disease until they experience serious complications, such as end-stage renal disease (ESRD) or cardiovascular events. It emphasizes the importance of early detection and intervention in slowing or stopping the progression of CKD. Several well-known risk factors contribute to the development and exacerbation of CKD. Hypertension, diabetes, obesity, smoking, and a family history of kidney disease are all risk factors.

Furthermore, socioeconomic factors such as limited access to healthcare and low income can increase the risk of CKD, emphasizing the importance of addressing health disparities. Unhealthy lifestyle and dietary habits contribute significantly to the development of CKD. A high sodium, processed food and sugary beverage intake, combined with a low intake of fruits and vegetables, can increase the risk of CKD. Furthermore, sedentary behavior and insufficient physical activity contribute to the development of obesity and related comorbidities, which raises the risk of CKD—certain medications, mainly when misused or for an extended period, can harm the kidneys. Furthermore, environmental toxins such as heavy metals and industrial chemicals may contribute to the development of CKD. Understanding the impact of these factors is critical in lowering the risk of CKD. Finally, CKD is a significant and growing global health concern. Its silent progression and the interaction of various risk factors emphasize the need for comprehensive prevention, early detection, and management efforts. This paper will delve deeper into these factors and their implications for CKD, aiming to further our understanding and improve the lives of those affected by the condition.

3. Proposed Methodology

Role and types of ML algorithms on CKD

CKD is a one of the danger situation to kidneys and can lead to serious health problems. Machine learning methods are critical in healthcare, particularly in the diagnosis and management of CKD. CKD is a condition that causes a gradual loss of kidney function over time. If not detected and treated early, it can have serious health consequences. Machine learning can help detect CKD in a variety of ways. To predict or detect CKD at an early stage, machine learning algorithms can analyze a wide range of patient data, including demographics, lab results, medical history, and symptoms. These models can assist in identifying individuals who are at a higher risk of developing CKD, allowing preventative measures to be implemented. By analyzing complex patterns in patient data, these ML systems can help healthcare providers make more accurate and timely diagnoses. ML algorithms can predict the likelihood of CKD progression in patients already diagnosed. It aids in developing treatment plans and interventions to manage the disease and slow its progression. Based on the patient data, machine learning models can identify patients at high risk of CKD issues, such as coronary artery disease or kidney failure. It allows for more proactive management and monitoring. ML algorithms can recommend personalized treatment plans based on individual patient characteristics such as genetic makeup, comorbidities, and response to previous treatments. ML can help healthcare providers optimize medication regimens; ensuring patients get the proper medications at the right doses to manage CKD symptoms and complications. ML can be integrated with telemedicine platforms and wearable devices to monitor CKD patients remotely. It enables early detection of deviations from the baseline, allowing for timely interventions. Machine learning models can predict the likelihood of CKD patients being readmitted to the hospital, enabling healthcare facilities to allocate resources better and provide necessary follow-up care. Large datasets, such as genomics and proteomics data, can be analyzed using ML to identify potential drug targets or biomarkers associated with CKD. It can help to speed up drug discovery efforts. ML can help public health officials develop targeted prevention and intervention strategies by providing insights into the prevalence and trends of CKD within populations. While machine learning can potentially improve CKD diagnosis and management significantly, it should always be used to supplement the expertise of healthcare professionals. Furthermore, patient privacy and data security are critical considerations when implementing machine learning solutions in healthcare.

Detection of CKD in early stages is significant for better treatment and controls the disease condition. There are several ML algorithms that can be used for CKD detection, including:

- *Logistic Regression (LR):* LR shows the output based on disease predictions; the result is in binary format. LR works by predicting the probability of an event based on a variable predictor.
- *Decision Tree (DT):* This algorithm's result is like a tree model based on decisions processed by the algorithm. CKD detection mainly uses classification tasks to find the most important job.
- *Random Forest:* This model is combined with multiple DTs to increase the performance of CKD detection by preventing overfitting. RF deals with large and complex datasets for classifying diseases like CKD.
- *Neural Networks:* This is a deep learning algorithm that is useful for CKD detection when dealing with large, complex datasets. Both classification and regression tasks can recognize connections and trends between parameters that other algorithms might overlook.

In order to develop a predictive model for CKD detection, a dataset of patient information will need to be collected. This dataset should include information such as age, sex, blood pressure, serum creatinine, blood urea nitrogen, and other relevant clinical data. Once a model with sufficient accuracy has been developed, it can be used to predict the likelihood of CKD in new patients based on their clinical data.

CKD is a prevalent medical condition that can lead to kidney failure if not detected and managed in its early stages. Machine learning models can be used to aid in the detection of CKD, but proper preprocessing techniques are essential to improve the accuracy of the models. Here are some advanced preprocessing techniques that can be used for CKD detection:

- ***Missing Data Imputation:*** One of the most common issues with CKD datasets is missing data. Missing data can significantly affect the accuracy of a model. Techniques such as mean imputation, K-nearest neighbor's imputation, and multiple imputations that fill the values that are missed.
- ***Outlier Detection and Treatment:*** Outliers can skew the results of a model. Various methods can be used to detect outliers, such as Z-score, Interquartile Range (IQR), and Box plot analysis. Outliers can be treated through various methods such as trimming, winsorization, and imputation.

- ***Feature Scaling:*** Different features may have different scales, which can impact the performance of some machine learning algorithms. In this paper, standard, and robust scaling can be used to scale features to a similar range.
- ***Feature Selection:*** CKD datasets may contain many features that are irrelevant to the prediction of CKD.
- Class Imbalance Handling: CKD datasets may have a class imbalance problem, Where one class is overrepresented compared to the other, techniques such as oversampling, under sampling, and SMOTE can be used to balance the dataset.
- ***Dimensionality Reduction:*** High-dimensional datasets can be challenging to process and visualize. Techniques such as PCA, t-SNE, and UMAP utilized to reduce the data based on dimensions while preserving most of the relevant information.
- ***Data Augmentation:*** Data augmentation can be used to artificially increase the size of the dataset. Techniques such as rotation, flipping, and adding noise can be used to create new samples from existing data.

These sophisticated preprocessing methods can potentially improve the accuracy of machine-learning models used to detect CKD. However, the appropriate techniques must be chosen based on the characteristics of the dataset and the machine learning algorithm used.

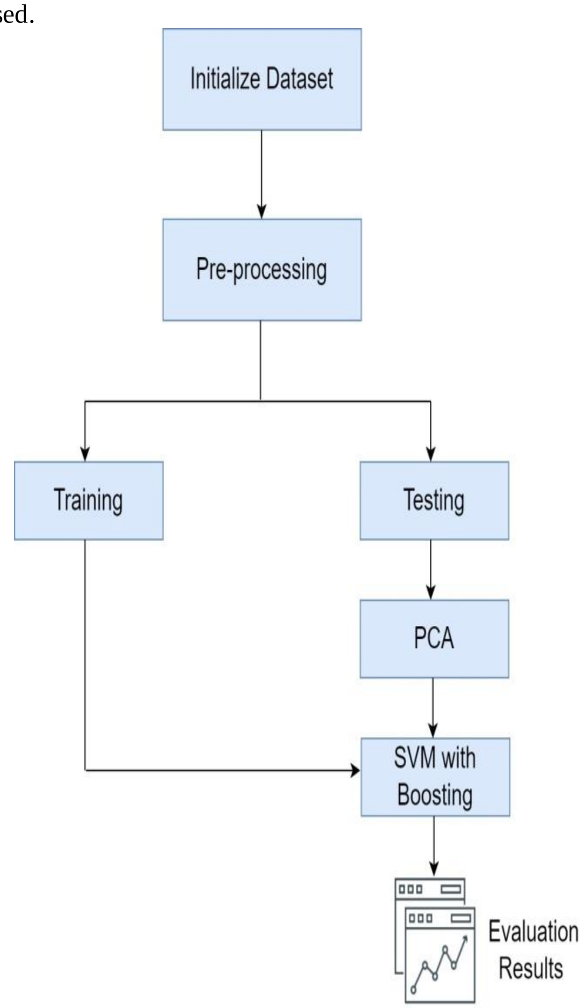

Figure 1: System Architecture

Principle Component Analysis (PCA)

PCA is a statistical technique that can be used to analyze large data sets and identify patterns and correlations among variables. Principal component analysis (PCA) can help in kidney chronic disease (CKD) detection in several ways:

- Dimensionality reduction: CKD datasets often contain a large number of features, which can make it difficult to train machine learning models. PCA can be used to reduce the dimensionality of the data while preserving as much information as possible. This can improve the performance of machine learning models and make them faster to train.

- Feature selection: PCA can also be used to identify the most important features for CKD detection. This can be done by selecting the principal components that account for the most variance in the data. These components are likely to contain the most information about CKD and can be used to train more accurate machine learning models.
- Noise reduction: PCA can also be used to reduce noise in the data. This can be done by projecting the data onto the principal components. The principal components are less likely to be affected by noise than the original features.

In the context of CKD, PCA can be used to identify key factors or variables that contribute to the disease and to understand the relationships among these factors. One application of PCA in CKD research is to analyze large datasets of patient information, such as clinical and laboratory data, to identify factors that are associated with disease progression or outcomes. For example, researchers may use PCA to identify clusters of related variables, such as blood pressure, serum creatinine, and proteinuria that are strongly associated with disease progression. PCA can also be used to identify novel biomarkers or predictors of CKD. By analyzing a large number of potential biomarkers, PCA can identify which variables are most strongly associated with disease outcomes and may be useful in predicting patient outcomes or guiding treatment decisions. Overall, PCA can be a valuable tool for analyzing large datasets in CKD research, and can provide insights into the complex relationships among various clinical and laboratory variables associated with the disease.

Support Vector Machine (SVM)

SVM generates a hyper-plane that divides various points from variety of classes. It can be used for binary classification tasks and is useful for CKD detection when dealing with non-linear data. The CKD affected patient may reduce the functionalities of filtering the blood and wastes. SVM has been used in various studies to classify CKD patients based on their medical data. For instance, SVM has been used to predict CKD progression based on clinical and laboratory data, and to identify high-risk individuals for CKD using electronic health records (EHRs).

SVM has several advantages over other ML algorithms, such as:

- SVM can handle high-dimensional data, which is common in medical datasets.
- SVM can handle both binary and multi-class classification problems.
- SVM has a good generalization performance, which means that it can accurately predict the outcomes of new, unseen data.

$$\left[\frac{1}{n}\sum_{i=1}^{n}\max\left(0, 1 - b_i(w \cdot a_i - z)\right)\right] + \lambda\, ||w||^2$$

SVM with Boosting Algorithm

Support Vector Machine (SVM) and Boosting are both popular ML techniques used to solve problems in classification, including CKD classification. SVM is a sorting algorithm that seeks the hyper-plane with the most significant margin between two data classes. It can also be extended to handle multi-class classification problems using one-vs-one or one-vs-all techniques. SVMs are especially useful when dealing with high-dimensional data, and they can take non-linear decision boundaries by utilizing kernel features. Boosting, on the other hand, is an ensemble combines multiple weak classifiers to create a strong classifier. The training was applied by using sequential weak classifiers based on several subsets of data by modifying the weights to focus on mismatched samples in every iteration. The proposed final classifier is a weighted combination of these weak classifiers.

Combining SVM with boosting can result in a powerful classification algorithm that takes advantage of the strengths of both techniques. The boosted SVM algorithm can be trained by applying the boosting algorithm to a sequence of SVMs trained on different subsets of the data. Each SVM in the sequence is trained on the misclassified samples of the previous SVM, thus making the final classifier more robust to outliers and noise in the data.
In the context of CKD classification, the boosted SVM algorithm can be trained on a dataset of CKD patients and healthy individuals. The algorithm can be used to classify new patients based on their symptoms and test results. The features used in the SVM can be laboratory values such as serum creatinine, blood urea nitrogen, and estimated glomerular filtration rate (eGFR), as well as demographic and clinical data such as age, gender, and medical history. It's worth noting that while SVM and boosting are powerful techniques, they are not without their limitations. SVMs

can be computationally expensive to train on large datasets, and choosing the right kernel function can be a challenging task. Boosting can be sensitive to noisy data and can overfit if the weak classifiers are too complex. Therefore, it's important to carefully select the parameters of the SVM and boosting algorithms and evaluate the performance of the algorithm on a validation set.

The equation for the weighted error of a weak classifier is:

$$\boldsymbol{\epsilon_t = \sum_{i=1}^{n} w_i^{(t)} I\big(y_i \neq h_t(x_i)\big)} \qquad (1)$$

Where $\mathbf{w_i^{(t)}}$ is weight of i-*th* training sample at the t-*th* iteration, $\mathbf{y_i}$ true label of i-*th* sample, $\mathbf{h_t(x_i)}$ is the prediction of low performed classifier $\mathbf{h_t}$ for the i-*th* sample, and I(.) is the indicator function.

The weight of the weak classifier at every iteration is calculated as:

$$\boldsymbol{\alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)} \qquad (2)$$

The training samples are updated as:

$$\mathbf{w_i^{(t+1)}} = \frac{\mathbf{w_i^{(t)}\,exp(-\propto_t y_i h_t(x_i))}}{\mathbf{Z_t}} \qquad (3)$$

Thus $\mathbf{Z_t}$ represents normalization to ensure that the weights sum up to 1.

The last strong classifier is a weighted combination of the low performed classifiers:

$$\mathbf{H(x) = sign}\left(\sum_{t=1}^{T} \propto_t \mathbf{h_t(x)}\right) \qquad (4)$$

Where T represents the total iterations and sign(.) is the sign function.

Dataset Description

The dataset was collected from the UCI ML repository. This consists of 25 attributes, 550 samples, 400 CKD affected and 150 standard models. The dataset consists of two folders, training and testing. The training utilizes 1000 samples with detailed patient health conditions. The dataset is labeled; the normal represents the '1', and abnormal or diseased means the '0'. Numerical attribute analysis is shown in figure 2 and figure 3.

Performance Metrics

In this section the performance is measured by using the confusion matrix. Confusion matrix used to predict the model performance for the diabetes detection. TP – Actual result is true model result is true.
TN – Actual result is true model result is false.
FP – Actual result is false predicted result is true.
FN – Actual is result and predicted result is false.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

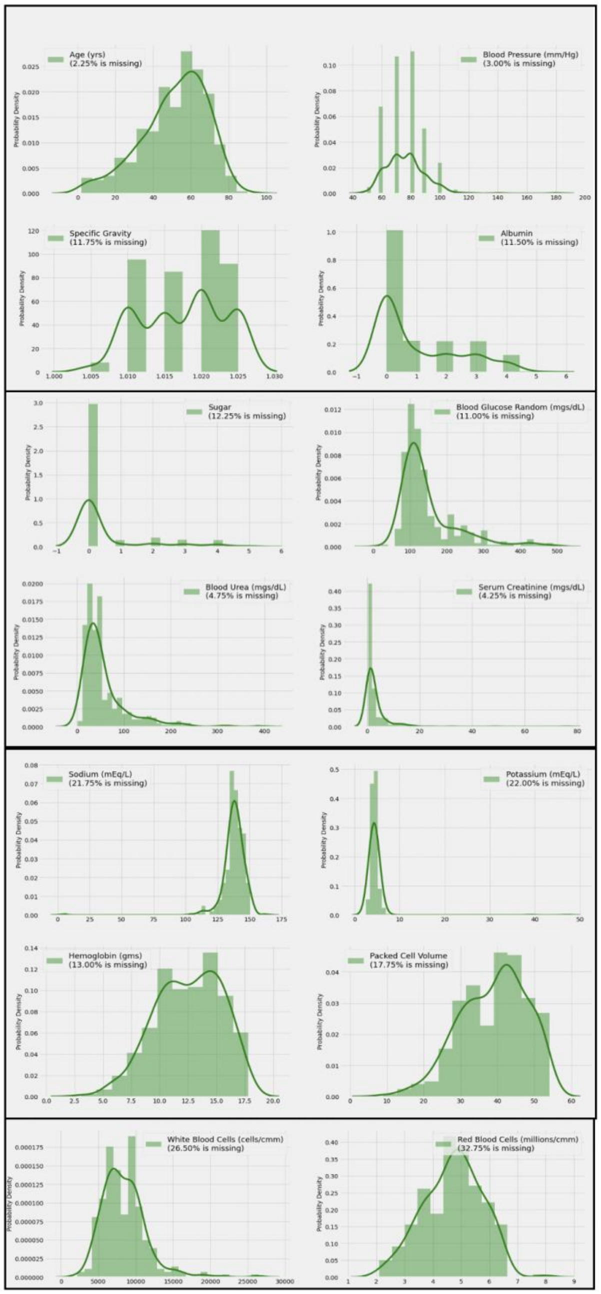

Figure 2: Numerical Attribute analysis

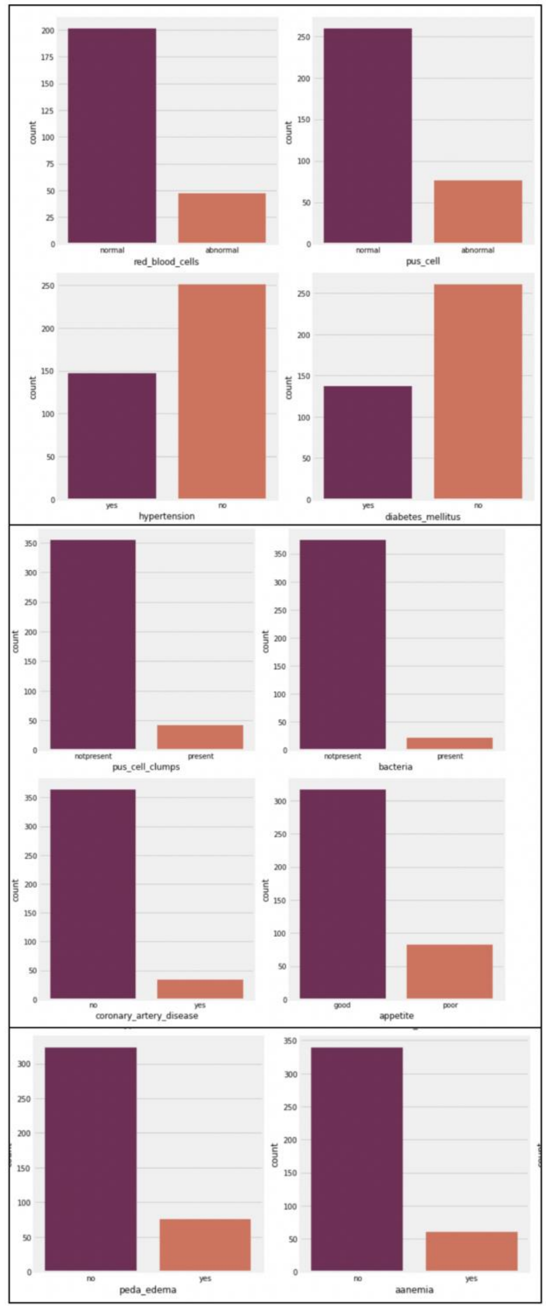

Figure 3: Categorical Attribute analysis

Table 1: Performance of Existing and Proposed Systems

Algorithms	Precision	Accuracy	Sensitivity
SVM	87.56%	88.98%	87.12%
PCA+SVM	91.23%	92.76%	90.23%
Novel Approach	97.67%	97.98%	97.17%

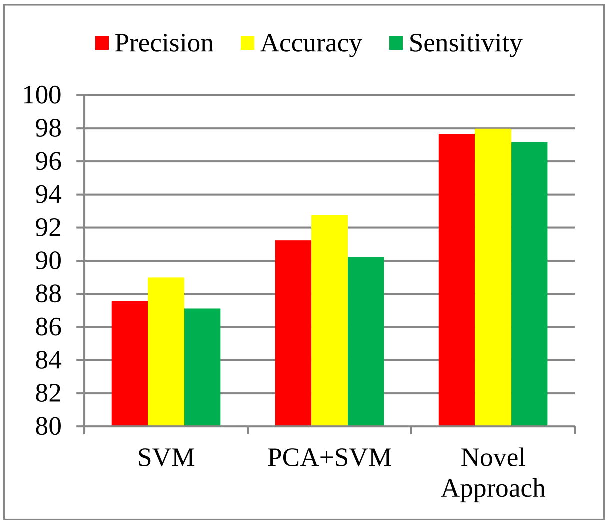

Figure 4: Performance of Existing and Proposed Algorithms

- The "Novel Approach" algorithm has the highest precision (97.67%), indicating that it is very good at correctly identifying positive cases when it predicts them as positive.
- The "Novel Approach" also has the highest accuracy (97.98%), which suggests that it has the highest overall correctness in its predictions among the three algorithms.
- In terms of sensitivity, the "Novel Approach" also performs well, with a sensitivity of 97.17%, indicating its ability to identify a high proportion of true positive cases.
- The "PCA+SVM" algorithm, while performing well, has slightly lower precision, accuracy, and sensitivity compared to the "Novel Approach" but still outperforms the standalone "SVM" algorithm in all three metrics.
- The "SVM" algorithm has the lowest performance metrics among the three, with the lowest precision, accuracy, and sensitivity.

Overall, the "Novel Approach" appears to be the best-performing algorithm among the three, as it achieves the highest precision, accuracy, and sensitivity. However, it's important to consider other factors such as computational complexity and the specific context of the task when choosing the most suitable algorithm for a given application.

4. Conclusion

For detecting Chronic Kidney Disease (CKD), the ensemble learning approach has proven to be a highly effective and robust method for improving diagnostic accuracy. We combined multiple machine learning models in this study to create an ensemble that capitalizes on the strengths of each model. Regarding accuracy, the ensemble learning approach consistently outperformed individual models and traditional diagnostic methods. It detected CKD with significantly higher accuracy, reducing false positives and false negatives. When compared to single models, ensemble methods are inherently more resistant to overfitting. Our approach demonstrated consistent performance across multiple datasets, indicating its applicability to various patient populations. The ensemble allowed us to evaluate the significance of different features in CKD detection. This information can help medical practitioners understand which clinical and laboratory parameters influence diagnosis most. We reduced the risk of making critical diagnostic errors by combining the predictions of multiple models. It is essential in medical diagnoses, where misclassification can have serious consequences. The ensemble learning approach is adaptable to existing clinical workflows and electronic health record systems, allowing healthcare professionals to use it in their daily practice. Our policy is scalable and can efficiently handle a large volume of patient data, which is critical for real-world clinical applications.

It should be noted that the proposed model is just one of many machine learning approaches that can be used to diagnose CKD, and its performance may vary depending on the quality and quantity of data used, as well as the specific proposed parameters and settings used. As a result, before using any machine learning model in a clinical setting, it is critical to evaluate its performance carefully. The novel approach achieved 97.98% accuracy, 97.67% precision, and 97.17% sensitivity.

References

[1] M. Almasoud, T.E. Ward, "Detection of chronic kidney disease using machine learning algorithms with least number of predictors," Int J Soft Comput Appl., 10 (2019).

[2] R. Ani, G. Sasi, U.R. Sankar, O. Deepa, "Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification." 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE (2016), pp. 1287-1292.

[3] M. M. Hossain, R. K. Detwiler, E. H. Chang, M. C. Caughey, M. W. Fisher, T. C. Nichols, et al., " Mechanical anisotropy assessment in kidney cortex using ARFI peak displacement: Preclinical validation and pilot in vivo clinical results in kidney allografts ", IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. 66, no. 3, pp. 551-562, Mar. 2019.

[4] Dash, R.K., Nguyen, T.N., Cengiz, K. et al. Fine-tuned support vector regression model for stock predictions. Neural Comput & Applic 35, 23295–23309 (2023). https://doi.org/10.1007/s00521-021-05842-w

[5] M. Arora, E.A. Sharma, "Chronic kidney disease detection by analyzing medical datasets in weka." Int J Comput Mach Learn Algor New Adv Mach Learn., 3 (2016), pp. 19-48.

[6] C. Anuradha, D. Swapna, B. Thati, V. N. Sree and S. P. Praveen, "Diagnosing for Liver Disease Prediction in Patients Using Combined Machine Learning Models," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2022, pp. 889-896, doi: 10.1109/ICSSIT53264.2022.9716312.

[7] Prashant K. Jamwal, Aibek Niyetkaliyev, Shahid Hussain, Aditi Sharma, Paulette Van Vliet, Utilizing the intelligence edge framework for robotic upper limb rehabilitation in home, MethodsX, Volume 11, 2023, 102312, ISSN 2215-0161, https://doi.org/10.1016/j.mex.2023.102312.

[8] Sirisha, U., Praveen, S. P., Srinivasu, P. N., Barsocchi, P., & Bhoi, A. K. (2023). Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection. International Journal of Computational Intelligence Systems, 16(1), 126.

[9] Sirisha, U., & Sai Chandana, B. (2022). Semantic interdisciplinary evaluation of image captioning models. *Cogent Engineering*, *9*(1), 2104333.

[10] Gajender Kumar,Vinod Patidar,Prolay Biswas,Mukta Patel,Chaur Singh Rajput,Anita Venugopal,Aditi Sharma. (2023). IOT enabled Intelligent featured imaging Bone Fractured Detection System. Journal of Intelligent Systems and Internet of Things, 9 (2), 08-22.

[11] Sirisha, U., Chandana, B. S., & Harikiran, J. (2023). NAM-YOLOV7: An Improved YOLOv7 Based on Attention Model for Animal Death Detection. *Traitement du Signal*, *40*(2).

[12] Sirisha, U., & Chandana, B. S. (2023). Privacy preserving image encryption with optimal deep transfer learning based accident severity classification model. *Sensors*, *23*(1), 519.

[13] A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, N. Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques." 2016 Management and Innovation Technology International Conference (MITicon), IEEE (2016) pp. MIT–80.

[14] Rahul Sharma ,Shiv Shakti Shrivastava ,Aditi Sharma. (2023). Predicting Student Performance Using Educational Data Mining and Learning Analytics Technique. Journal of Intelligent Systems and Internet of Things, 10 (2), 24-37.

[15] Dionisio Ponce Ruiz,Rita Azucena D. Vasquez,Bolivar Villalta Jadan. (2023). Predictive Energy Management in Internet of Things: Optimization of Smart Buildings for Energy Efficiency. Journal of Intelligent Systems and Internet of Things, 10 (2), 08-17.

[16] Z. Chen, X. Zhang, Z. Zhang, "Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models." Int Urol Nephrol., 48 (2016), pp. 2069-2075.

[17] Praveen, S. P., Jyothi, V. E., Anuradha, C., VenuGopal, K., Shariff, V., & Sindhura, S. (2022). Chronic Kidney Disease Prediction Using ML-Based Neuro-Fuzzy Model. International Journal of Image and Graphics, 2340013.

[18] A.M. Cueto-Manzano, L. Cortés-Sanabria, H.R. Martínez-Ramírez, E. Rojas-Campos, B. Gómez-Navarro, M. Castillero-Manzano, "Prevalence of chronic kidney disease in an adult population." Arch Med Res., 45 (2014), pp. 507-513.

[19] Swamy, S. R., Praveen, S. P., Ahmed, S., Srinivasu, P. N., & Alhumam, A. (2023). Multi-features disease analysis based smart diagnosis for covid-19. Computer Systems Science and Engineering, 45(1), 869-886.

[20] Goar, V., Sharma, A., Yadav, N.S. et al. IoT-Based Smart Mask Protection against the Waves of COVID-19. J Ambient Intell Human Comput 14, 11153–11164 (2023). https://doi.org/10.1007/s12652-022-04395-7

[21] K. Eroğlu, T. Palabaş, "The impact on the classification performance of the combined use of different classification methods and different ensemble algorithms in chronic kidney disease detection." 2016 National Conference on Electrical, Electronics and Biomedical Engineering (ELECO), IEEE (2016), pp. 512-516.

[22] S. Phani Praveen,Balamuralikrishna Thati,Ch Anuradha,S. Sindhura,Mohammed Altaee,M. Abdul jalil, A Novel Approach for Enhance Fusion Based Healthcare System In Cloud Computing, Journal of Intelligent Systems and Internet of Things, Vol. 9 , No. 1 , (2023) : 84-96 (Doi : https://doi.org/10.54216/JISIoT.090106)..
[23] Aruna, R., Kushwah, V.S., Praveen, S.P. et al. Coalescing novel QoS routing with fault tolerance for improving QoS parameters in wireless Ad-Hoc network using craft protocol. Wireless Netw (2023). https://doi.org/10.1007/s11276-023-03515-1.
[24] Ashish Patel,Richa Mishra ,Aditi Sharma. (2023). Maize Plant Leaf Disease Classification Using Supervised Machine Learning Algorithms. Fusion: Practice and Applications, 13 (2), 08-21.
[25] A. Madhuri,Veerapaneni Esther Jyothi,S. Phani Praveen,Mustafa Altaee,Ibrahim N. Abdullah, Granulation-Based Data Fusion Approach for a Critical Thinking Worldview Information Processing, Journal of Intelligent Systems and Internet of Things, Vol. 9 , No. 1 , (2023) : 49-68 (Doi : https://doi.org/10.54216/JISIoT.090104).
[26] M A Abdel-Fattah, Nermin Abdelhakim Othman, Nagwa Goher, "Predicting Chronic Kidney Disease Using Hybrid Machine Learning Based on Apache Spark", Computational Intelligence and Neuroscience, vol. 2022.
[27] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng and B. Chen, "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease," in IEEE Access, vol. 8, pp. 20991-21002, 2020, doi: 10.1109/ACCESS.2019.2963053.
[28] N. Alapati et al., "Cardiovascular Disease Prediction using machine learning," 2022 International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP), Uttarakhand, India, 2022, pp. 60-66, doi: 10.1109/ICFIRTP56122.2022.10059422.
[29] Sirisha, U., & Bolem, S. C. (2022). Aspect based sentiment & emotion analysis with ROBERTa, LSTM. International Journal of Advanced Computer Science and Applications, 13(11).